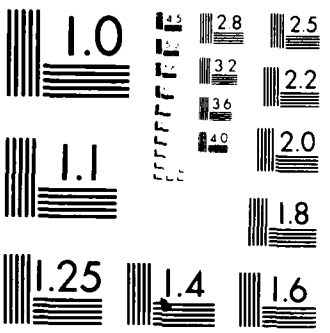
AD-A170 657 RESEARCH IN ADAPTIVE AND DECENTRALIZED STOCHASTIC CONTROL(U) TEXAS UNIV AT AUSTIN DEPT OF ELECTRICAL AND COMPUTER ENGINEERING S I MARCUS 17 MAY 85 1/1

UNCLASSIFIED AFOSR-TR-86-0549 AFOSR-84-0089 F/G 12/1 NL



AD-A170 657

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION: UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT: Approved for public release; distribution unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S): AFOSR-TR- 86-0549
6a. NAME OF PERFORMING ORGANIZATION: University of Texas at Austin; 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION: AFOSR/NM
6c. ADDRESS (City, State and ZIP Code): Dept. of Electrical & Computer Engineering, Austin, TX 78712	7b. ADDRESS (City, State and ZIP Code): Bldg. 410, Bolling AFB, DC 20332-6448
8a. NAME OF FUNDING/SPONSORING ORGANIZATION: AFOSR; 8b. OFFICE SYMBOL (If applicable): NM	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: AFOSR-84-0089
8c. ADDRESS (City, State and ZIP Code): Bldg. 410, Bolling AFB, DC 20332-6448	10. SOURCE OF FUNDING NOS. — PROGRAM ELEMENT NO.: 61102F; PROJECT NO.: 2304; TASK NO.: A1; WORK UNIT NO.:
11. TITLE (Include Security Classification): Interim Report: Research in Adaptive and Decentralized Stochastic Control (UNCLASSIFIED)	
12. PERSONAL AUTHOR(S): Marcus, Steven I.	
13a. TYPE OF REPORT: Interim; 13b. TIME COVERED: FROM _____ TO _____	14. DATE OF REPORT (Yr., Mo., Day): 1985 May; 15. PAGE COUNT: 11
16. SUPPLEMENTARY NOTATION	
17. COSATI CODES — FIELD / GROUP / SUB. GR.	18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number): Stochastic Adaptive Control, Decentralized Stochastic Control

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Significant progress was made in a number of aspects of stochastic systems. The problem of adaptive control of priority assignment in queueing systems was solved. A distance-measures approach to the problem of approximation and identification of queueing systems was studied. A problem of adaptively controlling a discounted-reward finite-state Markov decision process was solved. Major new results were obtained for the problem of adaptive control with incomplete observations. In particular, we have studied in depth a problem of adaptive control with incomplete observations, in which the state is a finite state Markov process.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT: UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☒ DTIC USERS ☐	21. ABSTRACT SECURITY CLASSIFICATION: UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL: Marc Jacobs	22b. TELEPHONE NUMBER (Include Area Code): (202) 767-4940; 22c. OFFICE SYMBOL: NM

DD FORM 1473, 83 APR EDITION OF 1 JAN 73 IS OBSOLETE.

Interim Report

March 15, 1984 - March 14, 1985

Grant AFOSR-84-0089

"Research in Adaptive and Decentralized Stochastic Control"

Steven I. Marcus
Department of Electrical and Computer Engineering
University of Texas at Austin
Austin, Texas 78712

May 17, 1985

Abstract

Significant progress was made in a number of aspects of stochastic systems. The problem of adaptive control of priority assignment in queueing systems was solved. A distance-measures approach to the problem of approximation and identification of queueing systems was studied. A problem of adaptively controlling a discounted-reward finite-state Markov decision process was solved. Major new results were obtained for the problem of adaptive control with incomplete observations. In particular, we have studied in depth a problem of adaptive control with incomplete observations, in which the state is a finite state Markov process.

I. SUMMARY OF RESEARCH PROGRESS AND RESULTS

During the first year of research supported by this grant, we have begun to make significant progress in a number of the areas which we proposed to investigate. In this section, we summarize the progress in those areas which have resulted in publications during the past year.

A. Adaptive Stochastic Control with Complete Observations

The assignment of priorities among customers (or demands or tasks) that arrive to a service station (or processor) is an important problem encountered in many situations, from computer networks to resource planning; the adaptive version of this problem is considered in [i]. In the priority assignment (or dynamic scheduling) problem, a single-server queueing system is considered whose customers are of K different classes. Customers of the several classes arrive according to independent Poisson processes with (known) mean arrival rates λ_i, i=1,...,K, and the service times, S_i, for class i customers are independent and identically distributed with unknown service rates $\mu_i = 1/m_i$, where $m_i = E(S_i)$. The state process is $X(t) = (X_1(t),...,X_K(t))$, where $X_i(t)$ is the number of class i customers in the system at time t, and the action space is $A = \{0,1,...,K\}$. The decision points $T_n (T_0 = 0)$ are the epochs at which either a service is completed or a customer arrives to find the server idle; if the action $a = ie A$ is chosen, then the next customer to be served is of class i, if $1 \leq i \leq K$, and $a = 0$ when the server chooses to be idle. A holding cost $c_i > 0$ is incurred for each unit of time that a class i customer stays in the system, so that a cost rate $k_1(x,a) = c_1 x_1 + ... + c_K x_K$ is incurred until the next transition occurs. Thus the expected cost is $c(x,a) = k_1(x,a)\tau(x,a)$.

Under the condition that the service times S_i have finite second

moment and that the total traffic intensity $\rho = \theta_1 m_1 + \ldots + \theta_K m_K$ satisfies the stability requirement that $\rho < 1$, it can be obtained that in the class of nonpreemptive work-conserving policies, an optimal stationary policy is the well-known "$c\theta$-rule" that ranks the classes so that $c_1\theta_1 \geq \ldots \geq c_K\theta_K$. Note that the $c\theta$-rule does not depend on the arrival rates.

It should be noted that, strictly speaking, the priority assignment problem is not included in the class of decision processes discussed above, because (under a stationary policy, like the $c\theta$-rule above) the process $X(t)$ is not semi-Markov; the process can have jumps (due to new arrivals) between two consecutive decision points. However, if we view a "transition" as taking place only at the decision points T_n defined above, namely, if instead of $X(t)$ we consider the process $X'(t) = X(T_n)$, $T_n \leq t < T_{n+1}$, then $X'(t)$ is semi-Markov. The important observation for our purposes, though, is that $X(t)$ itself is a semi-regenerative process with embedded Markov chain $X_n = X(T_n)$, $n=0,1,\ldots$, and that, under the stability assumption $\rho < 1$, the processes $X(t)$, $X'(t)$ and X_n have all the same limiting behavior, that is, the same limiting distribution. In summary, the moral is that our adaptive control scheme can be applied to more general problems provided that they can be reduced to equivalent semi-Markov decision problems.

With respect to the parameter estimation, we note that, since the unknown parameters $\theta_i = 1/m_i$ $(1 \leq i \leq K)$ are given in terms of the mean values m_i, the natural strongly consistent estimates to choose in Step II are $\hat{\theta}_{i,n} = 1/\hat{m}_{i,n}$, $n=1,2,\ldots$, where $\hat{m}_{i,n}$ are the sample mean (or first moment) estimates of the m_i. Their strong consistency follows from the law of large numbers, and from it we can immediately deduce Step III: as $n \to \infty$, $f(x,\hat{\theta}_n) \to f(x,\theta_0)$ a.s., for any state x, where f denotes the $c\theta$-rule, and $\hat{\theta}_n, \theta_0$ are the vectors of parameter estimates and true service rates, respectively.

Notice that, because of the particular form of this problem and the relationship between the observations and the unknown parameter, strongly consistent estimates are obtained from the easily computable sample mean; thus the modification of maximum likelihood proposed by Kumar and the strong hypotheses of other papers are unnecessary. Finally, Step IV, that is, the optimality of the adaptive $c\theta$-rule is verified in [i].

In [ii], we have considered general discounted-reward finite state Markov decision processes which depend on unknown parameters. An adaptive policy inspired by the nonstationary value iteration (NVI) scheme of Federgruen and Schweitzer is proposed; this is a variant of the usual method of successive approximations. It is shown that this adaptive policy is asymptotically discount optimal in the sense of Schäl. This NVI policy is compared with the certainty equivalent or naive feedback control (NFC) policy. The NFC requires computation and storage of the optimal policy for all values of the parameter θ; this represents considerable off-line computation and considerable storage, particularly if the parameter set is not finite. On the other hand, the NVI policy requires more on-line computation.

In related work, we have considered the identification and approximation of queueing systems in [iii]. In this paper, a distance-measures approach to such problems is taken. This approach combines ideas from statistical robustness, information-type measures, and parameter-continuity of stochastic processes. If one uses the appropriate distance measure, it is possible to obtain results on contiguity and asymptotic equivalence of the probability measures associated with the queueing systems, efficient estimates, most powerful tests, "quick" consistency, and other qualitative information that it would be difficult to obtain otherwise.

B. Adaptive Stochastic Control with Incomplete Observations

As we proposed, we have begun a major new direction of research involving adaptive estimation and control problems for stochastic systems with incomplete (or noisy) observations of the state. We have already been successful in obtaining some interesting new results; the first of these are reported in [iv]. In [iv], we consider discounted-reward, denumerable state space, Markov decision processes (MDP's) with incomplete state information and depending on unknown parameters. We are specifically interested in three problems: (a) How do we obtain a strongly consistent parameter estimation scheme based on partial state information? (b) How do we find "good" approximations of the optimal reward function? (c) How do we find (asymptotically) optimal policies, called below I-policies?

We approach these problems by following the usual procedure in which first the Markov decision process with incomplete state information (MDP-II) is transformed into a Markov decision process with complete state information (MDP-I) whose state space $\Phi := P(S)$ is the space of all probability measures on the state space S of the original MDP-II. Thus, since these two processes are equivalent -- in the sense that their optimal reward functions are equal -- problems (a), (b) and (c) are then transformed into the standard situation of a completely observed MDP-I with Polish (i.e., complete separable metric) state space Φ. Having done this, we can conclude the following: (i) There exists a sequence of estimators of the unknown parameters, which is strongly consistent for any I-policy. (ii) A nonstationary value-iteration (NVI) scheme can be used to solve both problems (b) and (c).

Part (i) is obtained by giving conditions on the MDP-II which imply the strong consistency of the conditional least squares estimators of Klimko and Nelson. To obtain (ii) we use the NVI scheme of Federgruen and Schweitzer

and the NVI adaptive policy [iv] to Markov decision processes with Polish state and action spaces. Thus, in short, we show that results for parameter-adaptive discounted MDP's with complete state observations [ii] under the usual (continuity and compactness) assumptions can be extended to partially observed MDP's with unknown parameters.

In [v], we have begun the investigation of the adaptive estimation and control of finite state Markov processes, as we proposed. The state is a finite state Markov chain $x_t \in \{\gamma_1, \ldots, \gamma_n\}$ with primitive transition matrix Q. The observation process $y_t \in \{0,1\}$. If Q is known, there is a finite dimensional recursive filter for $p_{t+1|t} = [p^1_{t+1|t}, \ldots, p^n_{t+1|t}]^T$, where $p_{t+1|t} = P[x_{t+1} = \gamma_i | y_0, \ldots, y_t]$:

$$p_{t+1|t} = Q^T p_{t|t-1} + (S^T p_{t|t-1} - Q^T \Sigma_t \gamma)[\gamma^T p_{t|t-1} - (\gamma^T p_{t|t-1})^2]^{-1}(y_t - \gamma^T p_{t|t-1}) \quad (1)$$

where $\Sigma_t = p_{t|t-1} p^T_{t|t-1}$. If x_{t+1} and y_t are conditionally independent given x_t, then $S = \Gamma Q$, where $\Gamma = \text{diag}(\gamma_1, \ldots, \gamma_n)$, and (1) can be rewritten in the following useful ways:

$$p_{t+1|t} = \frac{Q^T(I-\Gamma)}{1-\gamma^T p_{t|t-1}} p_{t|t-1} + \frac{Q^T[\Gamma - (\gamma^T p_{t|t-1})I]}{\gamma^T p_{t|t-1}(1-\gamma^T p_{t|t-1})} p_{t|t-1} y_t \quad (2)$$

$$= \frac{Q^T(I-\Gamma)}{1-\gamma^T p_{t|t-1}} p_{t|t-1}(1-y_t) + \frac{Q^T\Gamma}{\gamma^T p_{t|t-1}} p_{t|t-1} y_t \quad (3)$$

In general, the adaptive estimation problem involves the computation of estimates (e.g., state estimates) in the presence of unknown parameters; in addition, estimates of the parameters are often computed simultaneously. In the present context, the adaptive estimation problem is that of computing recursive estimates of the conditional probability vector when the transition matrix Q is not completely known (i.e., it depends on a vector of unknown parameters θ -- henceforth, we express this dependence via $Q(\theta)$). The

approach to this problem which we investigate in [v] has been widely used in linear filtering: we use the previously derived recursive filter for the conditional probabilities, and we simultaneously recursively estimate the parameters, plugging the parameter estimates into the filter. For example, for the filter (3), the adaptive filter would have the form:

$$\epsilon_t = y_t - \gamma^T \bar{p}_{t|t-1} \tag{4}$$

$$\hat{\theta}_t = \hat{\theta}_{t-1} + \alpha_t R_t^{-1} \psi_t \epsilon_t \tag{5}$$

$$\bar{p}_{t+1|t} = \frac{Q(\hat{\theta}_t)^T(I-\Gamma)}{1-\gamma^T \bar{p}_{t|t-1}} \bar{p}_{t|t-1}(1-y_t) + \frac{Q(\hat{\theta}_t)^T \Gamma}{\gamma^T \bar{p}_{t|t-1}} \bar{p}_{t|t-1} y_t \tag{6}$$

where $\{\alpha_t\}$ is a sequence of positive scalars, R_t is a positive definite matrix which modifies the search direction, and $-\psi_t$ is an approximation of the gradient of ϵ_t with respect to θ (evaluated at $\hat{\theta}_{t-1}$). We take R_t to be given by the Gauss-Newton direction:

$$R_t = R_{t-1} + \alpha_t [\psi_t \psi_t^T - R_{t-1}]. \tag{7}$$

Also, ψ_t is obtained by deriving an equation for $\partial \epsilon_t(\theta)/\partial\theta$ (for a fixed θ), and then evaluating at $\theta = \hat{\theta}_t$; thus

$$-\psi_t = \partial \epsilon_t / \partial \theta = -\gamma^T \partial p_{t|t-1}/\partial\theta$$
$$\stackrel{\Delta}{=} -\gamma^T \zeta(t). \tag{8}$$

Equations for $\zeta(t)$ (and for $\bar{\zeta}(t)$, obtained by substituting $\hat{\theta}_t$ for θ in the $\zeta(t)$ equations) are derived.

These computations give rise to a recursive stochastic algorithm of the general form

$$\eta_{k+1}^{\epsilon} = \eta_k^{\epsilon} + a_k^{\epsilon} G(\eta_k^{\epsilon}, \xi_k^{\epsilon}) \tag{9}$$

where $\eta_k^{\epsilon} = (\hat{\theta}_k, R_k)$, $\xi_k^{\epsilon} = (x_k, y_k, \bar{p}_{k|k-1}, \bar{\zeta}(k))$. We follow the approach of

Kushner to the Ordinary Differential Equation (ODE) Method of analyzing (9). That is, we define $t_k^\epsilon = \sum_{i=0}^{n-1} a_i^\epsilon$ and suppose that $t_n^\epsilon \to \infty$ as $n \to \infty$. Define the piecewise-constant interpolated process $\bar{\eta}^\epsilon(\cdot)$ by $\bar{\eta}^\epsilon(t) = \eta_k^\epsilon$ on $[t_k^\epsilon, t_{k+1}^\epsilon)$. The idea is to show weak convergence of the sequence $\{\bar{\eta}^\epsilon(\cdot)\}$ to the solution of an ODE, which can then be used to conclude properties (such as convergence as $t \to \infty$) of the parameter estimates $\hat{\theta}_t$. The essential assumption is that $\{\xi_k^\epsilon\}$ depends on $\{\eta_k^\epsilon\}$ in such a way that if $\eta_k = \eta$, a constant, then $\{\xi_k^\epsilon\}$ has a unique invariant (or stationary) measure. In [v], we show that it does indeed have a unique invariant measure.

II. PUBLICATIONS

[i] O. Hernandez-Lerma and S. I. Marcus, "Optimal Adaptive Control of Priority Assignment in Queueing Systems," *Systems and Control Letters*, Vol. 4, April 1984, pp. 65-72.

[ii] O. Hernandez-Lerma and S. I. Marcus, "Adaptive Control of Discounted Markov Decision Chains," to appear in *Journal of Optimization Theory and Applications*, June 1985.

[iii] O. Hernandez-Lerma and S. I. Marcus, "Identification and Approximation of Queueing Systems," *IEEE Transactions on Automatic Control*, Vol. AC-29, May 1984, pp. 472-474.

[iv] O. Hernandez-Lerma and S. I. Marcus, "Adaptive Control of Markov Processes with Incomplete State Information and Unknown Parameters," submitted to *Journal of Optimization Theory and Applications*.

[v] S. I. Marcus and A. Arapostathis, "Analysis of an Identification Algorithm Arising in the Adaptive Estimation of Markov Chains," submitted to the 24th IEEE Conference on Decision and Control.

III. PROFESSIONAL PERSONNEL ASSOCIATED WITH THE RESEARCH EFFORT

1. Steven I. Marcus, Principal Investigator
2. Hangju Cho, Research Assistant
3. Hong G. Lee, Research Assistant
4. Ian Walker, Research Assistant
5. Chang-Huan Liu, Postdoctoral Research Associate
6. Onesimo Hernandez-Lerma, Postdoctoral Research Associate
7. Aristotle Arapostathis, Associate Investigator

IV. PAPERS PRESENTED

1. S. I. Marcus, "Recent Developments in Nonlinear Estimation Theory," Distinguished Lecturer Series, Department of Electrical Engineering, University of Houston, University Park, Houston, Texas, April 9, 1984.

2. S. I. Marcus, "Optimal Adaptive Control of Queueing Systems," 2nd Istanbul Workshop on Large Scale Systems, June 24-27, 1984, Istanbul, Turkey.

3. S. I. Marcus and E. K. Westwood, "On Asymptotic Approximations for Some Nonlinear Filtering Problems," IFAC Triennial Congress, July 2-6, 1984, Budapest, Hungary.

4. J. W. Grizzle and S. I. Marcus, "A Jacobi-Liouville Theorem for Hamiltonian Control Systems," 23rd IEEE Conference on Decision and Control, December 12-14, 1984, Las Vegas, Nevada.

END

DTIC

9-86